Vehicle Classification Using a Biological Model of Hearing

D.A.Depireux and S.A.Shamma
Institute for Systems Research
University of Maryland, College Park, MD 20742-3311
didier,sas@isr.umd.edu
http://www.isr.umd.edu/CAAR

Work supported in part by MURI N00014-97-1-0501
from the Office of Naval Research
POC: Harold Hawkins
And by an ASEE fellowship to DAD.
POC for the present work:
Nassy Srour, Army Research Lab, Adelphi MD

Abstract :

The Army is interested in using acoustic sensors in the battlefield to perform vehicle tracking and classification using passive arrays of acoustic microphones and seismic sensors. Here, we present a prototype vehicle acoustic signal classification. To analyze acoustic features of the vehicle signal, we adopt biologically motivated feature extraction models. Physiological and psychophysical research have shown that primary auditory cortex performs to the first order a multi-scale decomposition of the incoming auditory spectra, on axes of log-frequency and time. This decomposition, based on the spectra emerging from a realistic model of the cochlea, is then used as a input to a classifier. Different vector quantization (VQ) clustering algorithms are implemented and tested for real world vehicle acoustic signal, such as Learning VQ, Tree-Structured VQ and Parallel TSVQ. Experiments on the Acoustic-seismic Classification Identification Data Set (ACIDS) database show that both PTSVQ and LVQ achieve high classification rates. The advantage of using biologically-based representation and classification algorithms include noise-robustness and existing low-power aVLSI implementations.
We present classication results and performance levels. The VQ schemes presented here have the advantage of not having to choose explicitly the features that distinguish one target from another. The burden is shifted to having to choose the ``best'' representation for the classifier.

Introduction :

A: The problem:

Our specific application involves a project of the Army Research Lab at Adelphi, MD. The Acoustic Detection System (ADS), to be integrated with the Surrogate Remote Sentry (SRS) system being developed by the Night Vision and Electronic Sensors Directorate (NVESD) of the Communications Electronics Command (CECOM), is an Advanced Technology Demonstration for the Rapid Force Projection Initiative.

Form SF298 Citation Data

Report Date *("DD MON YYYY")* 00002000	**Report Type** N/A	**Dates Covered (from... to)** *("DD MON YYYY")*

Title and Subtitle Vehicle Classification Using a Biological Model of Hearing	**Contract or Grant Number**
	Program Element Number
Authors Depireux, D. A.; Shamma, S. A.	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Institute for Systems Research University of Maryland College Park, MD 20742-3311	**Performing Organization Number(s)**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Monitoring Agency Acronym**
	Monitoring Agency Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Document Classification unclassified	**Classification of SF298** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** unlimited
Number of Pages 7	

The Acoustic Detection System consists in an omnidirectional acoustic sensor system that uses a circular array of microphones and a simple processor to detect, track, and classify targets in the battlefield. When targets are detected, the ADS determines lines of bearing (LOBs) to the targets relative to the position of the microphone array. When integrated with the SRS, the ADS can cue the SRS to approaching targets as they come within the detection range of the acoustic sensor. The ADS is developed at ARL for the long-range detection of ground and air vehicles in a typical battlefield environment. The current ADS' array of microphones is connected to a signal processing box that determines the targets frequency, signal-to-noise ratio, and classification of detected targets. The sensor array consists of seven sensors positioned in an hexagonal geometric configuration to maximize beamforming capabilities: the acoustic array is composed of ceramic microphones connected to an electronic box that contains signal conditioning amplifiers. A gain of 40 to 60 dB is provided to boost the signal levels so as to produce maximum voltage at a sound pressure level (SPL) of 120 dB.

The array is configured with seven microphones, six arranged in an 8-ft-diameter circle, and one in the center. The microphones are provided with 6-in.-diameter windscreens to reduce the effects of wind noise and are mounted on aluminum spikes that hold them vertically to the ground. The acoustic array can be configured in a variety of ways, depending on the mission. Currently, the beamformer (BF) software that resides in the processor unit can be reprogrammed to process LOB information based on a chosen array geometry. The acoustic system estimates target bearing using a frequency-domain BF to provide a degree of directional noise rejection and allow high-resolution estimation of the direction of arrival of the various signals.
The ADS is designed to detect, localize, and identify both air and ground combat vehicles. Narrow-band spectral analysis exploits the periodic nature of vehicular noise sources available in a typical battlefield environment. Subsequent operations process the spectral information to detect the sources of noise. A tracking function exploits the time continuity of the process to refine localization estimates and derive direction of travel and speed. Target bearing information is then used to cue an optical system.

Biologically inspired algorithms:

Why use biologically inspired algorithms?

The mammalian auditory pathway possesses remarkable abilities to detect, localize, separate and identify sounds even in the presence of noise or severe degradation. Studying the pathway provides us with a method to understand the nature and a representation of complex sounds. For a variety of systems, strategies imitating the structure of the peripheral auditory system have already been incorporated to analyze and transmit acoustic signals such as speech (e.g. the use of the bark scale in speech recognition systems). Our lab has experience in the design and building of a robotic head that can localize and point at a sound source in a noisy, reverberant room (in real time) with two degrees of freedom: azimuth (horizontal plane) and elevation (vertical plane).

How to use these algorithms ?

For the project under consideration, we are only concerned with azimuth detection, so the approach is from the biological and neurological perspectives of interaural time differences extraction.

In mammals, interaural time and level differences (ITD and ILD) provide cues for the angle of arrival. These are processed in the lateral and medial superior olives (LSO and MSO), which are located on the brainstem. Coincidence detection and inhibition are used to measure time and level differences, respectively, although the actual biological mechanism is not known, and we have considered two biologically plausible possibilities.

B: The algorithms:

Tree Structured Vector Quantizer:

TSVQ is an example of a classification tree where test vectors are classified stage by stage, with each stage giving a sharper classification than the previous. Each node of the tree is associated with a centroid, which can be thought of as a paradigm for a particular class. All test vectors start out by belonging to the root node. Then the vector is compared with the centroids of all nodes which are children of the node it currently belongs to. The vector is classified into the child with the centroid that is ``closest'' to it. The vector eventually ends up in a leaf node, and is assigned a class according to the class of the leaf node.

Making a vector quantizer (VQ) in the form of a tree offers several advantages. Firstly, if the tree is more or less balanced, the number of comparisons that have to be made are $O(\log n)$ where n is the number of partitions of the vector space. For a VQ on one level, we would have to make $O(n)$ comparisons. This improvement in search efficiency can be a big factor when we have a large number of classes. Also, as explained later, we can use parallel TSVQ techniques to further reduce the search time. Secondly, the way the underlying vector space is split at each node is frequently indicative of natural partitions in the data-set. The challenge is to preserve fidelity in the classification. Any substituting of an optimal partitioning of the signal vector-space by a tree structured partitioning reduces the optimality. Our goal is to make this difference as small as possible. Proper choice of pre-processing and tree-growing algorithm is crucial

C: Multi-Resolution TSVQ (MRTSVQ):

One special kind of VQ classifying tree is the Multi-Resolution TSVQ (MRTSVQ). Any particular test vector in an MRTSVQ is represented in multiple resolutions or scales. A method of creating such a representation is through affine wavelet transforms. Here, the auditory cortex filter used is an example of a multi-resolution transform. For a given vector we create a multi-resolution representation. At each level i in the tree, the ith resolution vector is compared against the centroids of the nodes in that particular level. The centroids at a particular level are also represented at the corresponding resolution. The vector is classified into the node that has the nearest neighbor centroid. At the next level, the next higher resolution of the vector is used for the comparison.

This method of classification offers one advantage over the unembellished TSVQ. At the higher levels, where more comparisons have to be made, we use a vector with fewer bits, thus doing many simple computations. Progressively finer details are added until satisfactory performance is obtained. This computational advantage is very important in online algorithms.

Cutting down on the data presented to the classifier in the early stages does not degrade performance much. In most cases of interest, one does not need all the available data to make the

relatively simple classification split that takes place at the higher levels. For example, in the case of a speaker ID, the decision whether the speaker is male or female can be made with a rather coarse representation of the sound.

We use a tree-growing algorithm that has been used by Baras and Wolk (Baras 1993) for classifying radar returns. The details of the algorithm can be found in the reference. The tree algorithm uses the Linde-Buzo-Gray (LBG) algorithm (Linde et al, 1980) for VQ at each level of the tree. The algorithm starts with an initially fixed number of centroids. The LBG algorithm is used to find a distribution of centroids that correspond to a local minimum in the expected squared error distortion. Then an additional centroid is introduced and LBG applied again to find the expected distortion. If the change in distortion from the addition is greater than a fixed fraction of the total distortion, another centroid is introduced and the process repeated. If the change in distortion is lesser than then fixed fraction, the algorithm goes to the next level. The cell in the current level is fixed and the leaf node with the highest value of the distortion is then split at the next lower level (higher resolution). This process goes on until a stopping criterion is satisfied. The stopping criterion can be a constraint on the final rate of the tree, the final number of leaf nodes, the final expected distortion of the tree, or any other criterion.

This is a greedy method of tree growing, in that the cell with the highest distortion in the current leaf nodes is the one that is split. There are other ways of choosing the split node, among which are, node with largest change in distortion for given rate increase, node with highest entropy and so on.

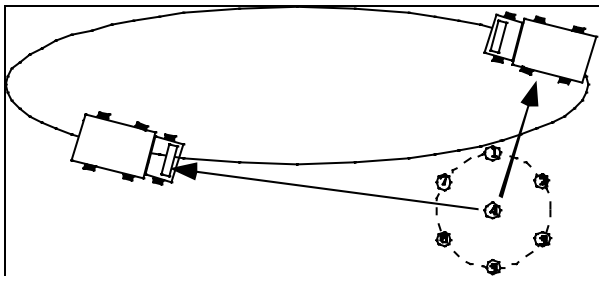

Figure 1 Schematic of the situation: two heavy vehicles are going around a track and the sound they generate is being recording by an array of microphones.

The algorithms

A.The auditory representation: A functional view of the auditory pathway and its attending repesentations has been presented before (Shamma et al, FedLab99). Many details are given there (and were already given in summary form at the IRIS 98 meeting) and will not be repeated there. The cochlea is viewed as a parallel bank of bandpass filters with a specific shape and constant Q-factor. The output of the cochlear filters forms an affine wavelet transform of the stimulus, with the log (frequency) spatial axis acting as a scale parameter. The cortical representation which

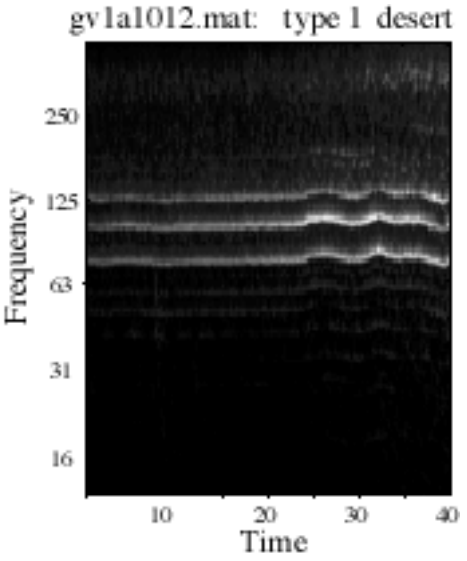

Figure 2: Cochlear pattern for ACIDS vehicle type 1, speed 5km/hr.

is used as a time-frequency decomposition preprocessing stage corresponds roughly to a time-frequency wavelet decomposition of the cochlear representation. Details can be found in Shamma et al 1989.

B.Training: In the ACIDS database, each vehicle has dozens of runs, corresponding to different speed and gear, different terrain (desert, arctic, normal roadway, and etc), and different recording systems. This database represents an ideal opportunity for classification research.

Type 1: heavy track vehicle
Type 2: heavy track vehicle
Type 3: heavy wheel vehicle
Type 4: light track vehicle
Type 5: heavy wheel vehicle
Type 6: light wheel vehicle
Type 7: light wheel vehicle
Type 8: heavy track
Type 9: heavy track

Table 1: Different vehicles in the ACIDS database.

Results:

In figure 2 we see a typical auditory time-frequency representation obtained by passing a vehicle signal through the cochlear filter banks. From this T-F representation, we find that the vehicle acoustic signal is between 20 to 200 Hz, and is dominated by salient low frequency harmonics parallel to the time axis. From this, we can derive the cortical representation that is used to classify vehicles.

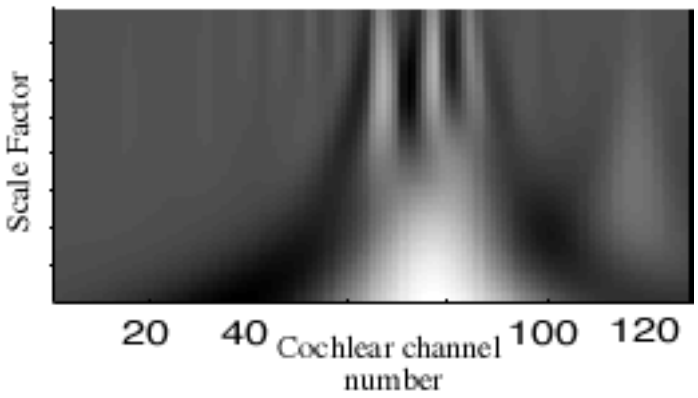

Figure 3: Multi-resolution representation from cortical representation

Fig. 3 shows a cortical processing pattern for the auditory spectrum. The coarse scale (lower part of the figure) captures the broad and skewed distribution of energy in the auditory spectrum, while the finer scale (upper part of the figure) captures the detailed harmonics structure. In the other intermediate cortical scales, the dominant harmonics are highlighted while the weaker ones are suppressed. Thus these intermediate scales emphasize the most valuable perceptual features within the signal. The cortical filter is a redundant representation, not all the scales are necessary for the classification algorithm.

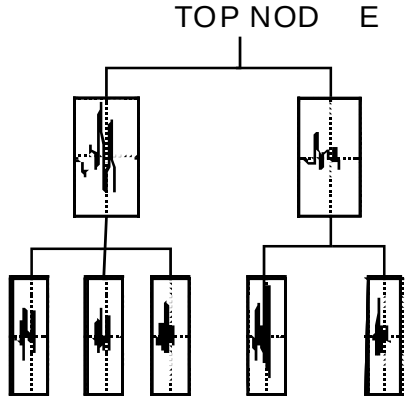

Figure 4: The first two levels of a VQ tree.

References [1]

1. J.S. Baras and S.I. Wolk, (1993) Hierarchical Wavelet Representations of Ship Radar Returns, Technical research report T.R. 93-100, Institute for Systems Research, U. of Maryland at College Park
2. Y. Linde, A. Buzo and R. Gray, An algorithm for Vector Quantizer Design, IEEE Trans. Comm., Vol COM-28, No. 1, pp 84-95, Jan 1980
3. S.A.Shamma, D.A.Depireux, C.P.Brown, N. Srour and T. Pham, Signal Processing in Battlefield Acoustic Sensor Arrays, in FedLab 99 proceedings

[1] The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

Predicted\True	1	2	3	4	5	6	7	8	9
1	**93.077**	37.5	18.958	2.2917	0	0	11.563	14.167	4.25
2	5.9615	**36.25**	2.7083	1.0417	0	0	3.125	0.2083	19
3	0.1923	0	**43.958**	0	0	0	5.3125	0	2.25
4	0	1.25	1.875	**78.542**	0	0	38.438	0.625	1.75
5	0	25	0.8333	0.625	**0**	0	1.25	0.4167	4.75
6	0	0	0.625	0	0	**0**	6.875	4.1667	6
7	0	0	4.5833	3.9583	0	0	**6.25**	0.2083	1.75
8	0.1923	0	3.125	3.5417	0	0	20.938	**79.583**	8.75
9	0.5769	0	23.333	10	0	0	6.25	0.625	**51.5**
Total %	100	100	100	100	100	100	100	100	100
Overall Score	46/71	Correct							

Table 2 Classification results from TSVQ as aplied to the ACIDS database